IDS 702: Module 3.3

Multinomial logistic regression

Dr. Olanrewaju Michael Akande

IDS 702

Recall logistic regression

- Recall that for logistic regression, we had

$$y_i|x_i \sim \text{Bernoulli}(\pi_i); \quad \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i$$

for each observation $i = 1, \ldots, n$.

- To get π_i, we solved the logit equation above to get

$$\pi_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

- Consider $Y = 0$ a baseline category. Suppose $\Pr[y_i = 1|x_i] = \pi_{i1}$ and $\Pr[y_i = 0|x_i] = \pi_{i0}$. Then, the logit expression is essentially

$$\log\left(\frac{\pi_{i1}}{\pi_{i0}}\right) = \beta_0 + \beta_1 x_i$$

- e^{β_1} is thus the (multiplicative) change in odds of $y = 1$ over the baseline $y = 0$ when increasing x by one unit.

IDS 702

Multinomial logistic regression

- Suppose we have a nominal-scale response variable Y with J categories. First, for the random component, we need a distribution to describe Y.

- A standard option for this is the multinomial distribution, which is essentially a generalization of the binomial distribution.
  Read about the multinomial distribution here and here.

- Multinomial distribution gives us a way to characterize

$$\Pr[y_i = 1] = \pi_1, \ Pr[y_i = 2] = \pi_2, \ \ldots, \ \Pr[y_i = J] = \pi_J, \quad \text{where} \ \sum_{j=1}^{J} \pi_j = 1.$$

- When there are no predictors, the best guess for each π_j is the sample proportion of cases with $y_i = j$, that is,

$$\hat{\pi}_j = \frac{\mathbf{1}[y_i = j]}{n}$$

- When we have predictors, then we want

$$\Pr[y_i = 1|\boldsymbol{x}_i] = \pi_{i1}, \ \Pr[y_i = 2|\boldsymbol{x}_i] = \pi_{i2}, \ \ldots, \ \Pr[y_i = J|\boldsymbol{x}_i] = \pi_{iJ}.$$

IDS 702

Multinomial logistic regression

- That is, we want the π_j's to be functions of the predictors, like in logistic regression.
- Turns out we can use the same link function, that is the logit function, if we set one of the levels as the baseline.
- Pick a baseline outcome level, say $Y = 1$.
- Then, the multinomial logistic regression is defined as a set of logistic regression models for each probability π_j, compared to the baseline, where $j \geq 2$. That is,

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \beta_{0j} + \beta_{1j}x_{i1} + \beta_{2j}x_{i2} + \ldots + \beta_{pj}x_{ip},$$

  where $j \geq 2$.
- We therefore have $J - 1$ separate logistic regressions in this setup.

IDS 702

Multinomial logistic regression

- The equation for each π_{ij} is given by

$$\pi_{ij} = \frac{e^{\beta_{0j}+\beta_{1j}x_{i1}+\beta_{2j}x_{i2}+\ldots+\beta_{pj}x_{ip}}}{1+\sum_{j=2}^{J} e^{\beta_{0j}+\beta_{1j}x_{i1}+\beta_{2j}x_{i2}+\ldots+\beta_{pj}x_{ip}}} \quad \text{for} \quad j > 1$$

and

$$\pi_{i1} = 1 - \sum_{j=2}^{J} \pi_{ij}$$

- Also, we can extract the log odds for comparing other pairs of the response categories j and $j^{\star}$, since

$$\begin{aligned}
\log\left(\frac{\pi_{ij}}{\pi_{ij^{\star}}}\right) &= \log\left(\pi_{ij}\right) - \log\left(\pi_{ij^{\star}}\right) \\
&= \log\left(\pi_{ij}\right) - \log\left(\pi_{i1}\right) - \log\left(\pi_{ij^{\star}}\right) + \log\left(\pi_{i1}\right) \\
&= \left[\log\left(\pi_{ij}\right) - \log\left(\pi_{i1}\right)\right] - \left[\log\left(\pi_{ij^{\star}}\right) - \log\left(\pi_{i1}\right)\right] \\
&= \log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) - \log\left(\frac{\pi_{ij^{\star}}}{\pi_{i1}}\right).
\end{aligned}$$

IDS 702

Multinomial logistic regression

- Each coefficient has to be interpreted relative to the baseline.
- That is, for a continuous predictor,
  - β_{1j} is the increase (or decrease) in the log-odds of $Y = j$ versus $Y = 1$ when increasing x_1 by one unit.
  - $e^{\beta_{1j}}$ is the multiplicative increase (or decrease) in the odds of $Y = j$ versus $Y = 1$ when increasing x_1 by one unit.
- Whereas, for a binary predictor,
  - β_{1j} is the log-odds of $Y = j$ versus $Y = 1$ for the group with $x_1 = 1$, compared to the group with $x_1 = 0$.
  - $e^{\beta_{1j}}$ is the odds of $Y = j$ versus $Y = 1$ for the group with $x_1 = 1$, compared to the group with $x_1 = 0$.
- Exponentiate confidence intervals from log-odds scale to get on the odds scale.

Significance tests

- For multinomial logistic regression, use the change in deviance test to compare models and test significance, just like we had for logistic regression.
- Fit model with and without some predictor x_k.
- Perform a change in deviance test to compare the two models.
- Interpret p-value as evidence about whether the coefficients excluded from the smaller model are equal to zero.

IDS 702

Model diagnostics

- Use binned residuals like in logistic regression.
- Each outcome level has its own raw residual. For each outcome level j,
  - make an indicator variable equal to one whenever $Y = j$ and equal to zero otherwise
  - compute the predicted probability that $Y = j$ for each record (using the `fitted` command)
  - compute the raw residual = indicator value - predicted probability
- For each outcome level, make bins of predictor values and plot average value of predictor versus the average raw residual. Look for patterns.
- We can still compute accuracy just like we did for the logistic regression.
- ROC on the other hand is not so straightforward; we can draw a different ROC curve for each level of the response variable. We can also draw pairwise ROC curves.

Implementation in R

- Install the package nnet from CRAN.
- Load the library: `library(nnet)`.
- The command for running the multinomial logistic regression in R looks like:

```
Modelfit <- multinom (response ~ x_1 + x_2 + ... + x_p, data = Data)
```

- Use `fitted(Modelfit)` to get predicted probabilities for observed cases.
- We will see an example in the next module.

IDS 702

What's next?

Move on to the readings for the next module!

IDS 702